

The proper study of mankind

Molecular biology has made the human genome accessible to laboratory investigation. But does that mean that the sequence of the human genome would be worth the effort?

THERE IS NO scientific reason for studying man. Thus one of the 120 speakers at this year's Cold Spring Harbor Symposium on the molecular biology of *Homo sapiens* from 28 May to 4 June. The issue arises because, as the programme was designed to demonstrate, there are now few fundamental questions in biology that cannot be explored using human genes and human cells.

This in itself is scarcely sufficient reason for overlooking the rather obvious shortcomings of man as an object of investigation. But there is one reason, just the same: the human species alone preserves its rare defective variants. Mouse mothers, whose matings are manipulable and whose generation times are manageable, tend to eat offspring that seem not quite right before they have grown large enough to be investigated, and thus destroy the rich mine of information that medical genetics has made available in man.

If we accept that man's sole unique contribution is his genetic diseases, the content of the Cold Spring Harbor Symposium has some interesting reflections to offer in the light of the debate, discussed in *Nature* (321, 371; 1986) and formally tabled at the symposium, on whether to launch into the gargantuan exercise of sequencing the entire human genome.

There have been two outstanding contributions of medical genetics to science. One is the analysis of mutant haemoglobins; the other is the remarkable achievement of Michael Brown and Joseph Goldstein (University of Texas, Dallas) who uncovered the cellular and molecular basis of the hypercholesterolaemia that predisposes to heart disease (see *Nature* 317, 569; 1985), the essence of which was reviewed by Goldstein at Cold Spring Harbor.

Cholesterol is cleared from the bloodstream in the form of a lipid-protein complex (LDL) that binds to a specific receptor (the LDL receptor) on the surface of cells, whence it is internalized by endocytosis and released to provide the substrate for the synthesis of cell membranes and steroid hormones, as appropriate. If the receptor-mediated internalization fails for any reason, the cholesterol builds up in the bloodstream, atherosclerotic plaques develop and heart disease ensues. By investigation of precisely why internalization does fail in individuals with familial hypercholesterolaemia, Brown

and Goldstein have built up a picture of the cellular and, more recently, the molecular biology of receptor-mediated endocytosis which has provided fundamental insights into these processes as well as clarifying the mechanism of atherosclerosis.

This has been possible only because Brown and Goldstein worked up to the molecular biology from classical genetics via biochemistry. More characteristic of latter-day triumphs were the papers presented by Louis Kunkel (Boston Children's Hospital) and Stuart Orkin (Harvard Medical School), both published in this issue of *Nature* (pp 73 and 32). They have arrived directly at the genes responsible for two human genetic diseases without engaging with the intervening biochemistry.

Kunkel and his collaborators have succeeded in forging a path to the gene for Duchenne muscular dystrophy (DMD); Orkin and his colleagues have identified, cloned and sequenced the gene for X-linked chronic granulomatous disease (CGD). Neither group is any the wiser about the molecular or cellular mechanisms of the disease.

From this point of view, DMD and CGD present problems of a different order. DMD is a programmed wasting disease of muscle whose mechanism and site of action are wholly unknown, whereas CGD is an immune deficiency known to be consequent upon a defect in the oxidase system in phagocytic cells. Peter Goodfellow explains in an article on p. 12 the route by which Kunkel and his collaborators have arrived at a region of 20,000 base pairs comprising possibly no more than a part of the DMD gene, and what this implies for the daunting prospect of identifying the gene product or products. The problem here, and much of the excitement, is that there is so little idea what to look for.

The CGD gene is much more tractable. Starting with genetic linkage analysis using polymorphic DNA probes, Orkin and co-workers identified a sequence that is transcribed specifically in normal phagocyte but not in the phagocytic cells of three of their four CGD patients. The fourth patient produced a transcript containing a small deletion that presumably abrogates the production of a functional protein. Whether the transcript encodes one of a number of known but poorly characterized components of the oxidase system, or some so far unknown enzyme,

may require some ingenuity, as well as some biochemistry, to discover.

The substantial challenge now facing Orkin and Kunkel and their collaborators is in no sense a reflection on the quality of the work or the validity of their approach. But it is an inevitable consequence of approaching the DNA directly, and thus has a direct bearing on the desirability of sequencing the rest of the human genome.

In the discussion on that topic at the symposium, Paul Berg (Stanford) set out the issues — is it feasible, who will pay and is it worth it? Walter Gilbert (Harvard) seraphically chalked up the tally — three thousand million bases at 10^5 bases per year equals 30,000 person-years or, at the current rate of 2×10^6 bases per year, but allowing for the 6 million already sequenced, 1,000–1,500 years to finish all of it; at a going rate of \$1.00 a base.

Nobody seems to doubt that it is feasible; nor that with improvements in technology it can be done in perhaps a century or less. The question is whether it is worthwhile. There is a general conviction that, even with funding from the US Department of Energy (see *Nature* 321, 371; 1986), the project would be bound to divert resources from other areas of research. In return, it is argued, it will encourage innovative technology and provide an information resource that would nourish other areas.

But technical innovations are already occurring with impressive momentum; and as an information resource, the sequence of the human genome is an extremely doubtful asset. If the skill and ingenuity of modern biology are already stretched to interpret sequences of known importance, such as those of the DMD and CGD genes, what possible use could be made of more sequences? Moreover, it is believed that roughly half the human genome is nonsense and repetition; and those who argue that there may be meaning in the nonsense and function in the repeats are not likely to prove their point by sequencing it all.

The difficulty is the same for all attempts to answer biological problems by reading DNA: unless you have a very good idea how to phrase the question, the sequence is not going to give you the answer. Blind sequencing at the expense of good ideas would, in the words of David Botstein, be for biologists to "indenture [themselves] to mindlessness".

Miranda Robertson